Ocean color model inversion using an artificial neural network

Steven A. Bailey
The Laboratory for Hydrospheric Processes, NASA, Goddard Space Flight Center, Wallops Flight Facility, Wallops Island, VA 23337, U.S.A.

Abstract. An artificial neural network (ANN) was trained to invert an ocean color surface spectral reflectance model based on two oceanic absorption constituents. These constituents are chlorophyll and chromophoric dissolved organic matter (CDOM). Limits for each constituent were set at 0.0 to 30.0 mg/m^3 for chlorophyll and 0.0 to 1.0 m^{-1} for CDOM. A four layer ANN was found to give good results. The final sum-squared error with 1000 random spectra of 31 channels each using vector length normalization was 9.5 x 10^{-2}.

1. Introduction

The objective of this paper is to demonstrate the use of an artificial neural network in inverting an ocean color model. This task is normally difficult due to the large number of independent variables available which can describe ocean color reflectance (Gordon *et al*. 1988). Simple methods, like color ratio algorithms seem to work best in case I waters, but fail in more complex case II waters (Doerffer and Fischer 1994). More complex methods like principal component analysis (Sathyendranath *et al*. 1989, Tassan 1994) can be used for case II waters, but have a large computational cost when resolving ocean color constituents.

Artificial neural networks have several distinct advantages compared to the above techniques. First, development of an inversion algorithm is not needed for the retrieval of ocean color constituents (Dawson *et al*. 1993). Second, after a neural network has been created, processing time for parameter retrieval is known and comparatively very short (Rumelhart and McClelland 1986). Third, neural networks have the ability to generalize in the face of noisy data (Wasserman 1989).

This study limited the number of ocean color constituents to chlorophyll and CDOM. Since chlorophyll is the main pigment of phytoplankton (Hoepffner and Sathyendranath 1993), other pigments were not considered. CDOM was chosen as the opposing component since current remote sensing techniques (Tassan 1994) retrieve CDOM only when chlorophyll concentrations remain low. This would be a good test of neural networks and their ability to cope with complex, non-linear relationships.

2. Reflectance model

The semianalytical radiance model of ocean color (Gordon *et al*. 1988, Hoge *et al*. 1995) was used to generate training data for the artificial neural network. Two constituents (chlorophyll and CDOM) were varied and applied to the model. Spectra from the model was input to the network and constituents were output. The goal was to teach the network to map spectra to constituents.

The following equation for water-leaving radiance $[L_w]_N$ was used. It is based on work by Gordon *et al*. (1988) and for this study modified by Hoge *et al*. (1995).

$$[L_w]_N = \left[\frac{(1-p)(1-\overline{p})F_0R}{m^2Q(1-rR)}\right] \qquad (1)$$

where p is the Fresnel reflectance of the sea surface for normal incidence; $\overline{p}$ is the Fresnel reflection albedo of the sea surface for irradiance from the sun and sky; m is the index of refraction of water; Q is the ratio of upwelling radiance to upwelling radiance toward the zenith; r is the water-air reflectance for totally diffuse irradiance; R is the irradiance reflectance just below the sea surface; and F_0 is the mean extraterrestrial solar irradiance. Model detail not discussed can be found in the references cited above.

A two-component model was chosen for three reasons. First, this model corresponds well with active-passive measurements made by AOL-POCS (Hoge *et al*. 1995). Second, training a neural network on fewer components reduces the training time and simplifies the network. Training is also more successful with fewer components. Finally, the active component of AOL-POCS used for constituent measurement does not distinguish CDOM from detritus, so detritus is not included in this study (Hoge *et al*. 1995).

Modeled spectra was based on a 256 channel CCD diode array. For training purposes, the number of channels was reduced to 31 spanning 403.3 nm to 722.5 nm. Normalizing all spectra and constituents was necessary for neural training. The best results were found after treating all spectra as vectors and normalizing each by vector length. Spectral shape remains unspoiled and spectral amplitude remains below one. Constituents are simply normalized from zero to one based on their minimum and maximum values.

3. Artificial neural networks

Artificial neural networks were developed in the mid 1950’s to mathematically mimic the first-order response of the nervous system (Wasserman 1989). The question was how to model the output of a hypothetical neuron given some number of inputs with varying signals applied to those inputs. This study continued until the late 1960’s when Minsky and Papert (1969) demonstrated that a single artificial neuron or perceptron

(Rosenblatt 1959) could be ‘trained’ to learn a linear function mapping these given inputs to an output. However, ‘mappings’ requiring non-linear relationships could not be represented by this single perceptron. It was believed that to accomplish non-linear mappings, more perceptrons were needed and they needed to be layered. This is where the output of one perceptron is the input to another percetptron. Since there was no known algorithm to train multi-layer networks, wide interest in neural networks subsided until the mid 1980’s.

Werbos (1974) discovered a technique involving vectors where one vector *R* could be represented by the dot product of many vectors *W*. He devised a mathematical method of altering those vectors *W* whereby they could represent not only vector *R*, but many other vectors *S* simultaneously. Vectors *W* then acted as a ‘feature detector’, parsing ‘features’ of vectors *S* that distinguished each from one another. This technique is known today as back-propagation (BP) of errors.

Rumelhart and McClelland (1986) rediscovered this technique (BP) and applied it to the study of artificial neural networks. They were then able to train multi-layer neural networks. With this technique came an explosion of interest across many fields because neural networks could now represent complex, non-linear relationships.

3.1. *Network construction*

For this study, a ‘multi-layer feed-forward’ artificial neural network was chosen. This network consists of an input layer, at least one hidden layer, and an output layer. Signals are fed to the input layer. From there, they propagate forward through any hidden layer(s) and then arrive at the output layer. During this propagation the mapping from an input signal or vector to the output vector takes place. The network used in this research is described in Figure 1.

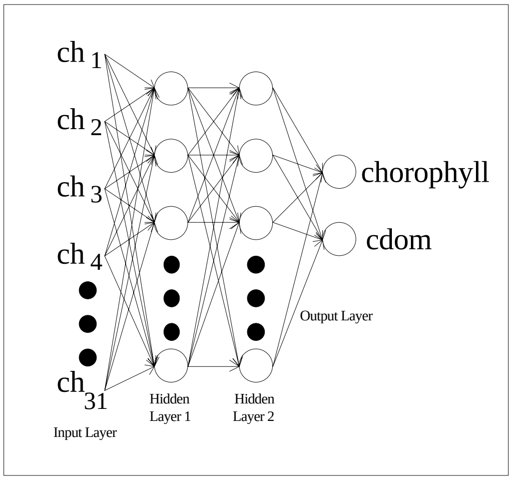

Figure 1. A four layer, feed-forward network. Lines denote weights and circles denote neurodes.

Each layer is connected by a series of weights. These weights are simply variables. Their collective purpose is to act as memory for the system. When a network learns to map an input vector to an output vector, this mapping is contained in the weights.

Hidden layers are those layers found between the input layer and the output layer. Hidden layers always contain neurodes. Neurodes are functions which transfer or squash (Wasserman 1989) the *Net*. The *Net* is the dot product of inputs (*x*) and weights (*w*) which attach to a particular neurode. A bias is usually added to the *Net*. Figure 2 is a representation of the *Net*.

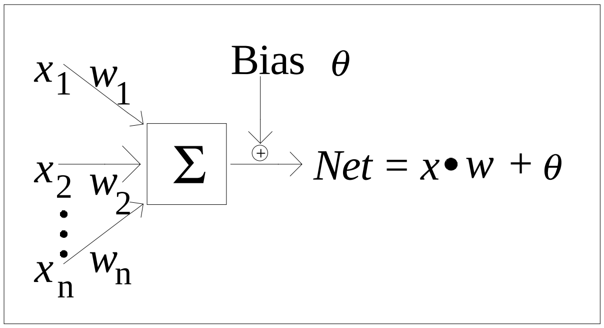

Figure 2. The *Net* of inputs, their associated weights plus a bias.

The transfer function most commonly used is the sigmoid (Figure 3). It has two noteworthy qualities. It is differentiable everywhere and exhibits a type of automatic gain control. When a signal is small (approaching zero), the gain of that signal when applied to the sigmoid is high due to a steep slope. When the signal is large, the sigmoid slope approaches zero. The signal is then attenuated.

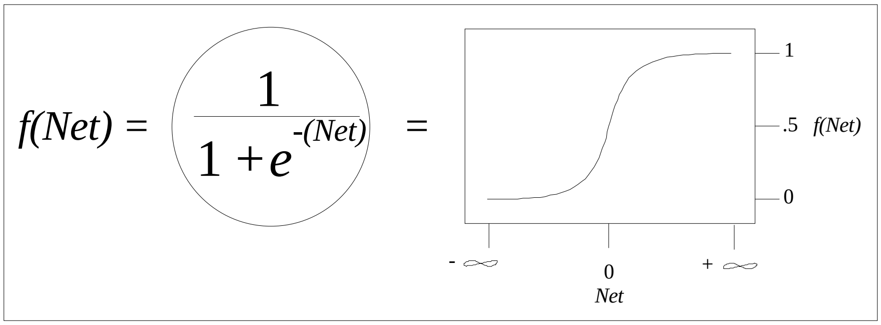

Figure 3. A typical transfer function used in an ANN.

3.2. *Back-propagation of errors*

Error is loosely defined as the difference between *what is known* versus *what is expected*. Since we are trying to train a neural network to learn some function, the error is calculated after each attempt to define this function. The objective is to minimize this error.

When training a multi-layer network, a systematic, mathematically based method is needed to ensure this error does diminish. As described in the previous section, the transfer function has one criteria. It must be differentiable everywhere (Wasserman 1989). One example of a function that meets this criteria is the sigmoid. It has the following form:

$$f\left(Net_i\right) = \frac{1}{1 + e^{-\left(Net_i\right)}} \tag{2}$$

where the derivative is:

$$f'\left(Net_i\right) = f\left(Net_i\right)\left(1 - f\left(Net_i\right)\right) \tag{3}$$

and *i* is a particular neurode.

For a three layer network, there is one input layer, one hidden layer, and one output layer. This means there are two levels of weights to adjust in an effort to reduce error. To calculate error for the output layer, one must take the difference between the desired response ($y_i^{Desired}$) and the actual response (y_i^{Actual}).

$$E_i^{Output\,layer} = y_i^{Desired} - y_i^{Actual} \tag{4}$$

Error for the hidden layer of a three layer network is:

$$E_i^{Hidden\,layer} = f'\left(Net_i\right)^{Hidden\,layer} \sum_{j=1}^{n} \left(W_{ij} E_j^{Output\,layer}\right) \tag{5}$$

where:

j = a particular neurode from the output layer.
n = the total number of neurodes from the output layer.
W_{ij} = a weight between the output layer and the hidden layer.

3.3 *Network training*

Training a neural network consists of applying some fixed number of patterns (in this case, spectra) along with their contributing constituents (chlorophyll and CDOM) to a network in an interative manner. After each spectrum is applied, an error is calculated for the output layer and any hidden layer, and an adjustment is made to all weights and biases. These adjustments are made via equations based on the delta rule (Caudill 1992).

$$\Delta W_i = \mathrm{B} E_i I_i + \alpha \, \Delta W_i^{\mathrm{Pr}\,evious} \tag{6}$$

$$\Delta \theta_i = \mathrm{B} E_i I_i + \alpha \, \Delta \theta_i^{\mathrm{Pr}\,evious} \tag{7}$$

where:

ΔW_i = Change in weight
$\Delta \theta_i$ = Change in bias
B = Learning constant, $0.0 < \mathrm{B} < 1.0$
E_i = Error
I_i = Input to neurode
α = Momentum, $0.0 < \alpha < 1.0$
$\Delta W_i^{\mathrm{Pr}\,evious}$ = Change in weight from previous interation
$\Delta \theta_i^{\mathrm{Pr}\,evious}$ = Change in bias from previous interation

Essentially, a particular weight or bias is changed by the amount of error found at a connecting neurode. The larger the error, the larger the change. Also, the above equations include a term for momentum. This term ensures training continues even when error gets stuck in a local minimum. Think of momentum as a ball rolling at some nominal velocity with just enough speed to roll out of any small depression it encounters.

After much experimentation, it was determined that a four layer artificial neural network worked best when inverting a two constituent, 31 channel ocean color model. A set of 1000 spectra were used to train the network. Initially, a step function was used to systematically cycle through the range of constituent values. However, results improved when those 1000 spectra where created by varying the two constituents randomly about their set limits. The limits for each constituent were:

Chlorophyll	0.1 - 30.0 mg/m^3
CDOM	0.01 - 1.0 m^{-1}

Not only must constituents be normalized from zero to one, but so must spectra. Several normalizations of spectra were tried, but the most promising was normalizing by vector length. If each spectra is a vector, then normalizing by vector length is found by the following:

$$S_n = \frac{S}{|S|} \tag{8}$$

where:

S_n = Normalized vector
S = Original vector
$|S|$ = Length or magnitude of vector $= \sqrt{S_1^2 + S_2^2 + \ldots + S_n^2}$

Training occurred by sequentially applying each of the 1000 spectra to the network. After each individual application, an error called the sum-squared error (SSE) was calculated. This is simply the sum of the squared differences between the desired response ($y_i^{Desired}$) and the actual response (y_i^{Actual}).

$$SSE = \sum_{i=1}^{Outputs} (y_i^{Desired} - y_i^{Actual})^2 \tag{9}$$

The sum of all the sum-squared errors is then calculated after all 1000 spectra have been applied. One iteration where all 1000 spectra have been sequentially trained by the network is called a cycle or epoch. The SSE for this study is then redefined as:

$$SSE = \sum_{i=1}^{1000} \sum_{j=1}^{Outputs} (y_{ij}^{Desired} - y_{ij}^{Actual})^2 \tag{10}$$

Figure 4 and Figure 5 show a small set of spectra which span the limits defined above. Spectra starting at number 0 have a chlorophyll concentration of 0.1 mg/m^3 and CDOM absorption of .01 m^{-1}. The chlorophyll concentration then increases in 0.3 mg/m^3 increments as CDOM increases in 0.01 m^{-1} increments. For example, spectrum 100 has a chlorophyll concentration of 30 mg/m^3 and CDOM absorption of 1 m^{-1}.

Figure 4 is based on spectra normalized by dividing all spectra by the peak value across all channels. Moving from small concentrations of constituents to larger concentrations of constituents creates a rapid drop in water leaving radiance. Saturation occurs and spectral shape changes very little.

Figure 5 is based on spectra normalized by vector length. There is more variability in spectra past number 10. Also, the dynamic range of spectra has been reduced. Both factors contributed to a better trained network. Therefore, our inversion provided better results.

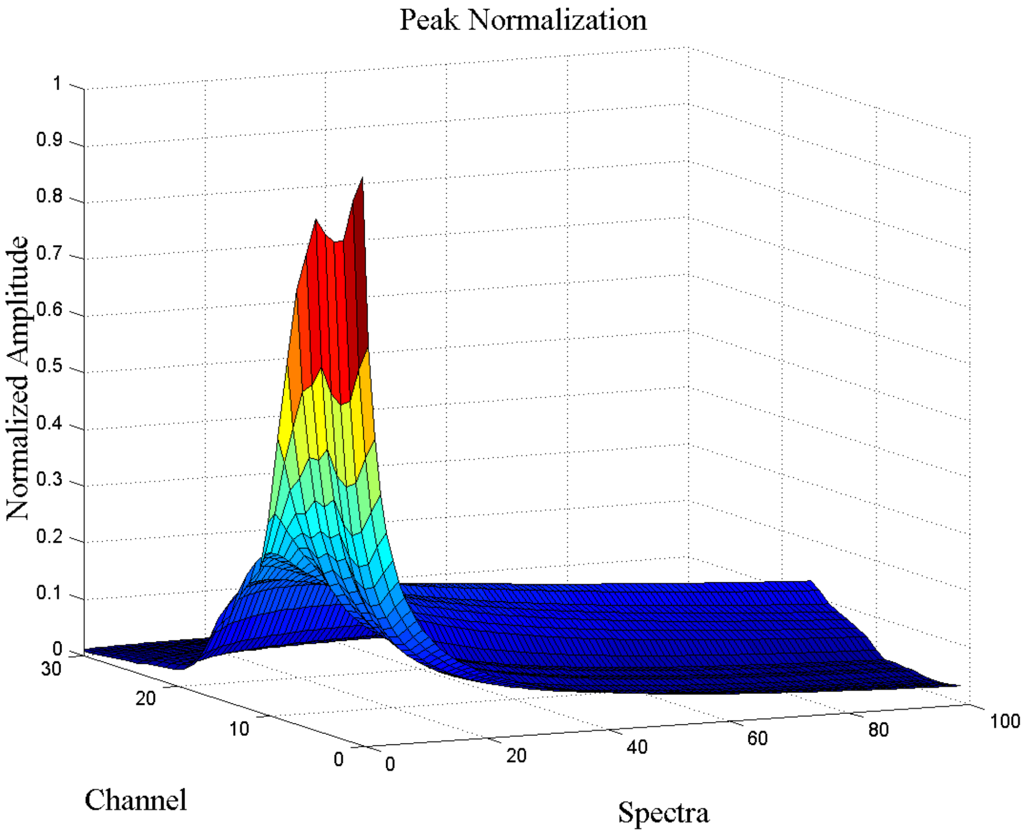

Figure 4. 100 spectra normalized by peak.

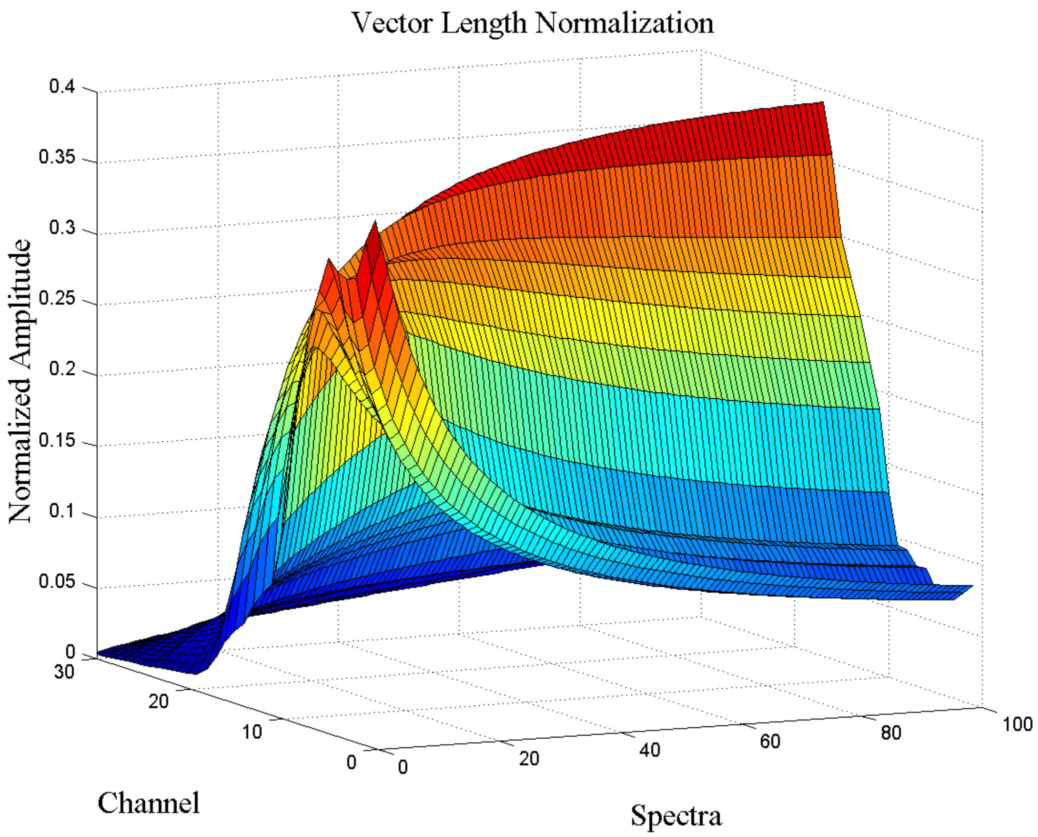

Figure 5. 100 spectra normalized by vector length.

A set of 1000 spectra were randomly generated based on the model and limits previously discussed. From these 1000 spectra, two training sets were generated based on normalizing by peak channel and normalizing by vector length. Each of these sets were then applied to the Stuttgart Neural Network Simulator (SNNS - version 4.1). The following SNNS specific parameters were used:

ANN model	= BackpropMomentum
Learning rate	= 0.2
Momentum	= 0.5
Flat spot elimination	= 0.01
Weight Initialization	= -1.0 to 1.0
Cycles or epochs	= 10000
Input layer size	= 31
Hidden layer 1 size	= 24
Hidden layer 2 size	= 16
Output layer size	= 2

Below are scatter plots for each method. The first two plots show chlorophyll and CDOM (known and calculated) based on a trained network from peak normalized spectra.

The next two plots show chlorophyll and CDOM (known and calculated) based on a different trained network using vector length normalized spectra. The data set normalized by peak channel achieved a SSE of 25.2 x 10^{-2}. The data set normalized by vector length achieved a SSE of 9.5 x 10^{-2}. Due to a lower SSE, a higher correlation is seen when using vector length normalized spectra.

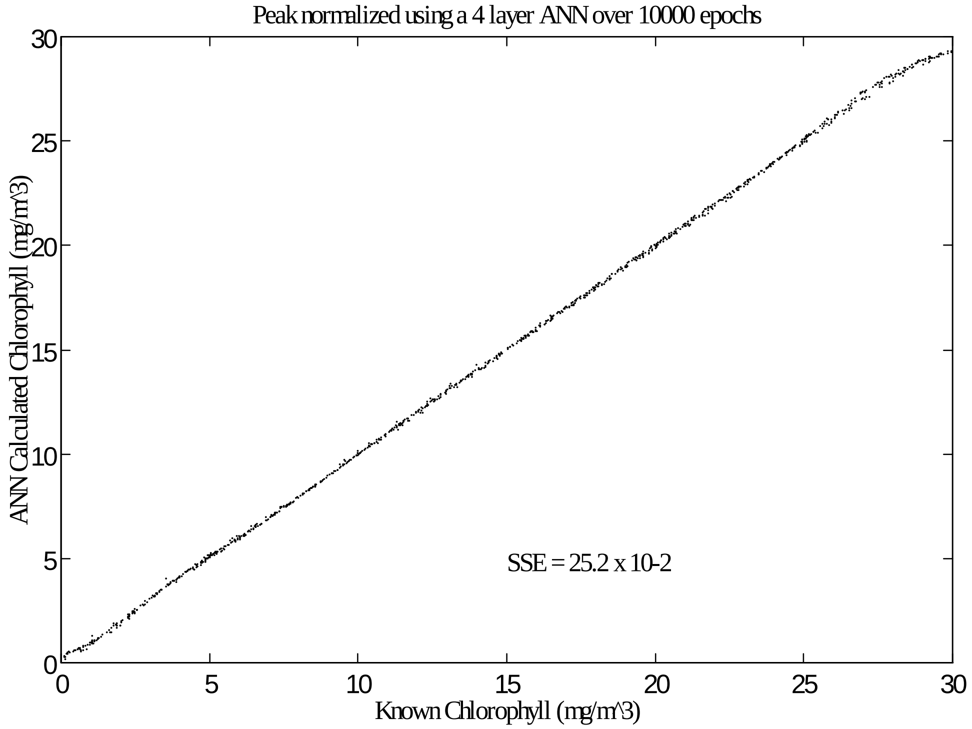

Figure 6. A scatter plot of known chlorophyll vs. ANN calculated chlorophyll using peak normalization.

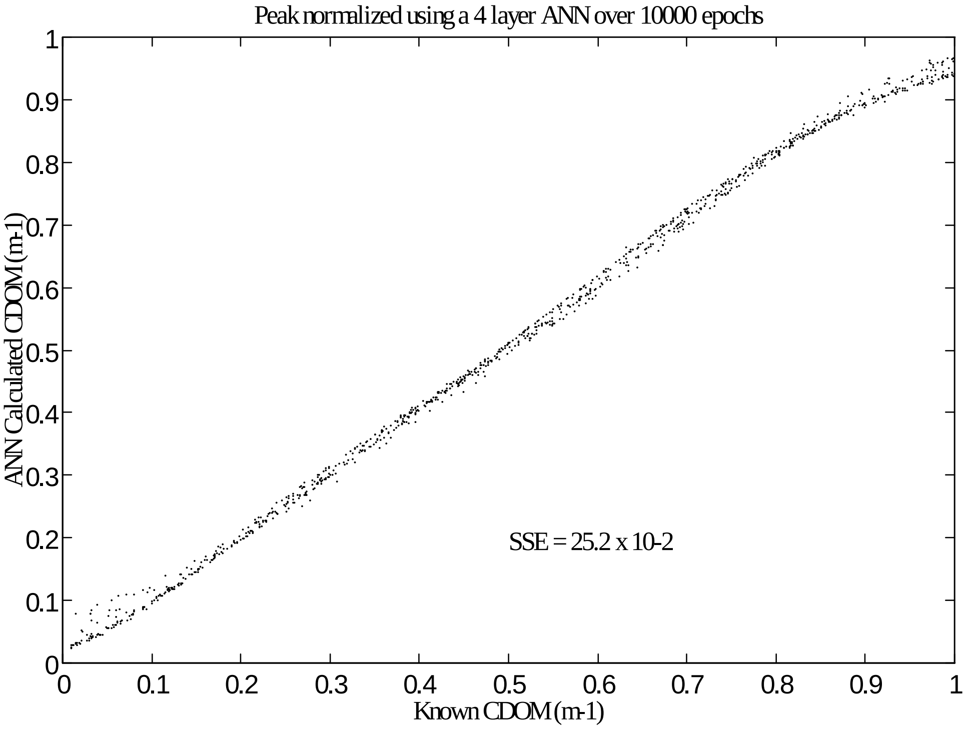

Figure 7. A scatter plot of known CDOM vs. ANN calculated CDOM using peak normalization.

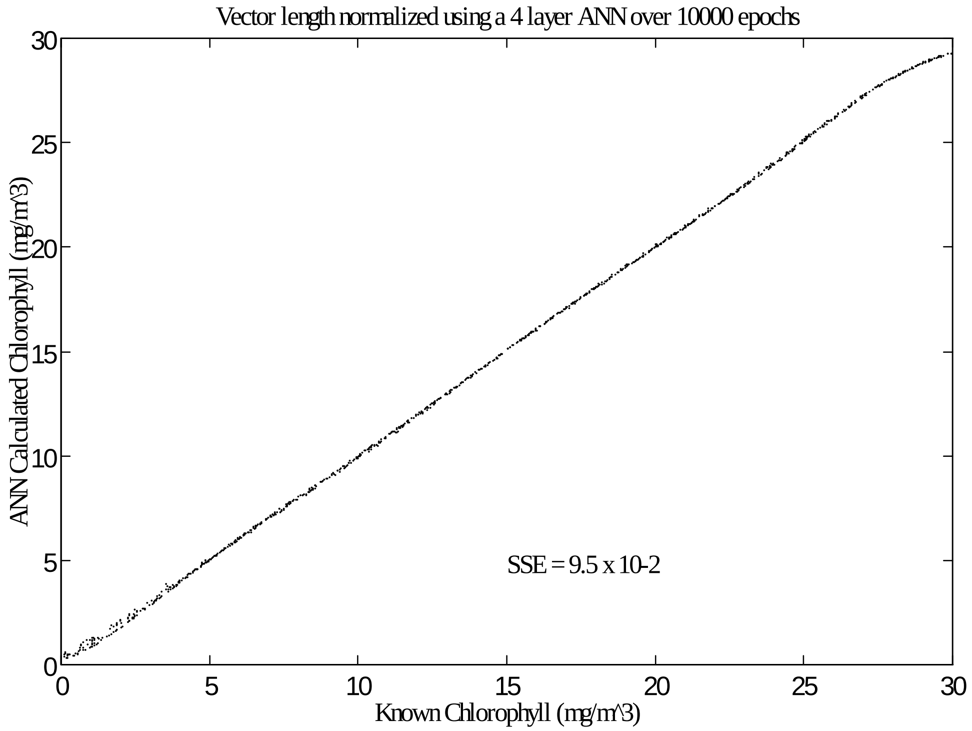

Figure 8. A scatter plot of known chlorophyll vs. ANN calculated chlorophyll using vector length normalization.

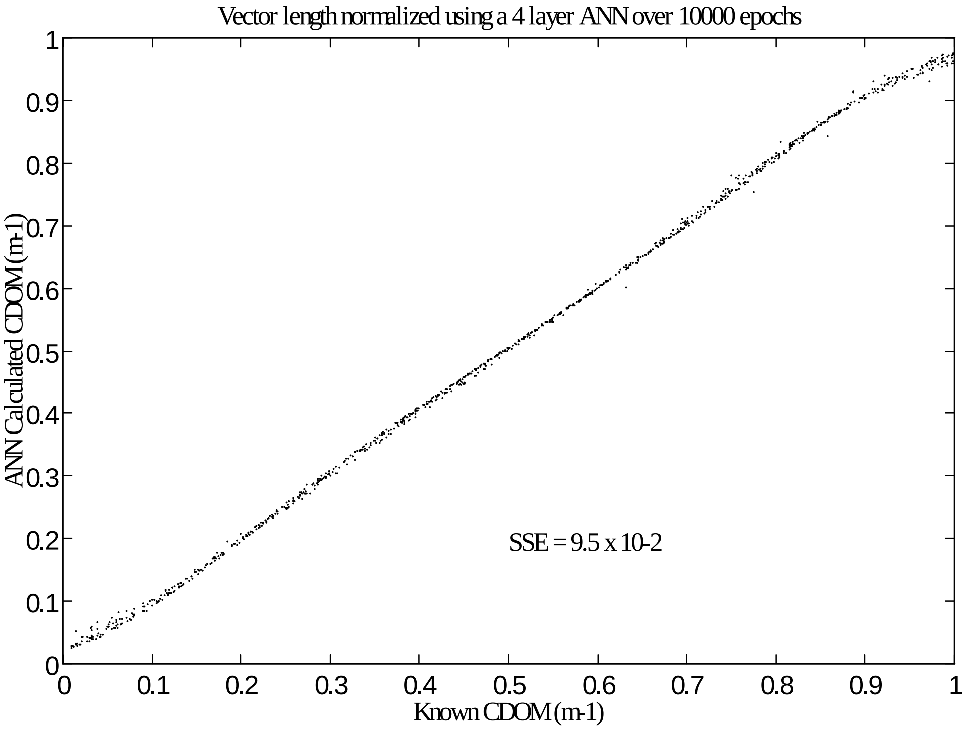

Figure 9. A scatter plot of known CDOM vs. ANN calculated CDOM using vector length normalization.

4. Conclusions

Parameter retrieval via artificial neural networks moves the burden of finding a valid algorithm via conventional means to creating a network that will train on some given data set. Training a network for parameter retrieval can become a fruitless endeavor if certain constraints are not considered. A data set that fully represents the parameter range desired must be generated or gathered. All data (input and output) must be normalized. Network structure must be deep enough to fully represent the data. An appropriate network paradigm must be chosen and training variables must be set so training converges to a solution. In all, much time can be spent determining these items.

Training a network on modeled ocean color data is a challenge. As constituents increase in concentration, spectral radiance decreases (Gordon *et al*. 1988). This leads to spectra which change very little in shape. For neural network training, this causes problems because of the large dynamic range in spectra. Training was found to improve by decreasing this dynamic range but leaving the spectra shape intact. Normalizing spectra by vector length achieves this goal.

It was shown that modeled ocean color spectra can be fully represented and inverted by an artificial neural network. The next goal is to apply real data to the currently trained network. It is then we will evaluate the validity of this technique.

Acknowledgments

The author thanks Dr F. Hoge of NASA for his continued interest and support with this project. Thanks also go to B. Swift, Dr M. Williams, and J. Yungel (all with EG&G) for their patience in answering all of my questions.

References

BISHOP, C. M., ROACH, C. M., 1992, Fast curve fitting using neural networks. *Review of Scientific Instruments*, 63, 4450-4456.

CAUDILL, M., BUTLER, C., 1992, *Understanding Neural Networks* (Cambridge, MA: MIT Press).

DAWSON, M. S., FUNG, A. K., 1993, Neural networks and their applications to parameter retrieval and classification. *IEEE Geoscience and Remote Sensing Society Newsletter*, 6-14.

DOERFFER, R., FISCHER, J., 1994, Concentrations of chlorophyll, suspended matter, and gelbstoff in case II waters derived from satellite coastal zone color scanner data with inverse modeling methods. *Journal of Geophysical Research*, 99, 7457-7466.

GORDON, H. R., BROWN, O. B., EVANS, R. H., BROWN, J. W., SMITH, R. C., BAKER, K. S., CLARK, D. K., 1988, A semianalytic radiance model of ocean color. *Journal of Geophysical Research,* **93**, 10909-10924.

HOEPFFNER, N., SATHYENDRANATH, S., 1993, Determination of the major groups of phytoplankton pigments from the absorption spectra of total particulate matter. *Journal of Geophysical Research,* **98**, 22,789-22,803.

HOGE, F. E., SWIFT, R. N., YUNGEL, J. K., 1995, Oceanic radiance model development and validation: application of airborne active-passive ocean color spectral measurements. *Applied Optics,* 34, 3468-3476.

HOGE, F. E., VODACEK, A., SWIFT, R. N., YUNGEL, J. K., BLOUGH, N. V., 1995, Inherent optical properties of the ocean: retrieval of the absorption coefficient of chromophoric dissolved organic matter from airborne laser spectral fluorescence measurements. *Applied Optics,* 34, 7032-7038.

MINSKY, M., PAPERT, S., 1969, *Perceptrons* (Cambridge, MA: MIT Press).

ROSENBLATT, R., 1959, *Principles of Neuradynamics* (New York, NY: Spartan Books).

RUMELHART, D. E., McCLELLAND, J. L., (Eds) 1986 , *Parallel Distributed Processing* (Cambridge, MA: MIT Press).

SATHYENDRANATH, S., PRIEUR, L., MOREL, A., 1989, A three-component model of ocean colour and its application to remote sensing of phytoplankton pigments in coastal waters. *International Journal of Remote Sensing,* **10**, 1373-1394.

TASSAN, S., 1994, Local algorithms using SeaWiFS data for the retrieval of phytoplankton, pigments, suspended sediment, and yellow substance in coastal waters. *Applied Optics*, **33**, 2369-2378.

WASSERMAN, P.D., 1989, *Neural Computing: Theory and Practice* (New York, NY: Van Nostrand Reinhold).

WERBOS, P. J., (Thesis) 1974, *Beyond Regression: New Tools For Prediction And Analysis In The Behavioral Sciences* (Harvard University, MA).